PHYSICAL REVIEW B **88**, 064302 (2013)

Nonlinear thermoelectricity in point contacts at pinch off: A catastrophe aids cooling

Robert S. Whitney

Laboratoire de Physique et Modélisation des Milieux Condensés (UMR 5493), Université Grenoble 1, Maison des Magistères, B.P. 166, 38042 Grenoble, France

(Received 30 August 2012; revised manuscript received 7 May 2013; published 8 August 2013)

I consider refrigeration and heat engine circuits based on the nonlinear thermoelectric response of point contacts at pinch off, allowing for electrostatic interaction effects. I show that a refrigerator can cool to much lower temperatures than predicted by the thermoelectric figure of merit ZT (which is based on linear-response arguments). The lowest achievable temperature has a discontinuity, called a *fold catastrophe* in mathematics, at a critical driving current $I = I_c$. For $I > I_c$ one can in principle cool to absolute zero, when for $I < I_c$ the lowest temperature is about half the ambient temperature. Heat backflow due to phonons and photons stops cooling at a temperature above absolute zero, and above a certain threshold turns the discontinuity into a sharp cusp. I also give a heuristic condition for when an arbitrary system's nonlinear response means that its ZT ceases to indicate (even qualitatively) the lowest temperature to which the system can refrigerate.

DOI: 10.1103/PhysRevB.88.064302 PACS number(s): 72.15.Jf, 84.60.Rb, 73.63.−b

I. INTRODUCTION

Nanostructures often have thermoelectric responses, with electrical currents causing heat currents, and vice versa.[1–3] There have recently been a number of proposals for nanostructures or molecules with large thermoelectric responses[4–11] which could have engineering applications for efficient thermoelectric power generation and refrigeration. In particular, it is hoped that they could cool electrons well below the temperature of standard cryostats,[12–15] which are increasingly inefficient at sub-kelvin temperatures.

However, *good* nanostructure refrigerators (those which cool to significantly below their environment's temperature) are rarely in the linear-response regime. Linear-response theory works for small temperature drops (compared with the average temperature) at the scale of the nanostructure and the scale of the electron's inelastic scattering length. This is often the case in bulk semoconductors,[16,17] but *not* in such nanostructures. See, for example, experiments on refrigeration with S-N tunnel junctions, that generate a temperature drop from 300 mK to 100 mK across a tunnel junction.[12–15]

Unfortunately, there is no general theory for the *nonlinear* response of quantum systems, because interaction effects are usually significant, and must be modeled using approximations appropriate for the system in question. Here, I calculate the *fully* nonlinear thermoelectric response of a point contact at pinch off. This system is one of the main candidates for a nanoscale thermoelectric, and its linear (and nearly linear) thermoelectric response is well studied experimentally[18,19] and theoretically.[18,20–22] I consider this thermoelectric response when the temperature drop across the point contact is of the order of the average temperature, for which the response is far outside its linear regime. This can be modeled with a nonlinear Landauer-Büttiker scattering theory[23–27] for thermoelectric heat transport.[28] I find that the dimensionless figure of merit, ZT, ceases to be a good measure of the thermoelectric response outside the linear regime. Electricity generation is *worse* than linear-response theory indicates, but refrigeration is *better* (achieving much lower temperatures than linear-response theory predicts). Indeed, the lowest temperature of the refrigerator is a discontinuous function of the electrical current. This discontinuity—a fold catastrophe in mathematical language—occurs at a critical current I_c, and helps refrigeration. For currents $I < I_c$ the refrigerator cannot cool below a finite temperature (about half the ambient temperature for $I \to I_c$), while for $I > I_c$ it passes the catastrophe and can *in principle* cool to absolute zero (see Fig. 1).

In practice a thermoelectric device's quality is reduced by the nonlinear backflow of heat carried by chargeless particles: phonons and photons. When such backflow effects are weak, the catastrophe is little affected, but cooling stops at a temperature above absolute zero. At a critical value of backflow effects, the catastrophe becomes a cusp (discontinuity in the derivative) of the dependence of the lowest temperature on I. The nonlinear nature of the cusp still means the lowest achievable temperature is lower than linear theory predicts.

II. NEARLY LINEAR ANALYSIS FOR ANY SYSTEM, AND ITS BREAKDOWN

The usual "nearly linear" analysis[1] takes linear response theory plus a joule heating term, and enables one to quantify devices in terms of their dimensionless figure of merit $ZT = G\Pi^2/[(\Theta_{el} + \Theta_{ph})T]$, where T is the device temperature, Π is its Peltier coefficient, G and Θ_{el} are electrical and thermal conductances of electrons, while Θ_{ph} is the thermal conductivity of chargeless excitations, principally phonons and photons. This nearly linear analysis predicts electric power generation (when the island is heated) with an efficiency

$$\eta = \frac{\sqrt{ZT+1}-1}{\sqrt{ZT+1}+1}\left(1-\frac{T_0}{T_{isl}}\right), \tag{1}$$

where T_{isl} and T_0 are the island and environment temperatures. Typically, ZT is taken at the temperature $\sim \frac{1}{2}(T_0 + T_{isl})$. Carnot efficiency corresponds to $ZT \to \infty$. For refrigeration, it predicts that the lowest achievable temperature, T_{min}, is given by

$$T_{min}/T_0 = 1 - \tfrac{1}{2}ZT. \tag{2}$$

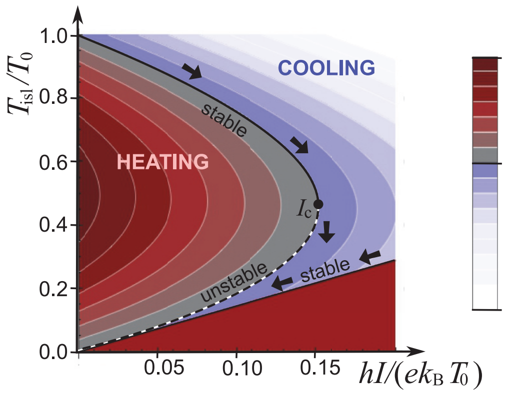

FIG. 1. (Color online) Heat current $J(T_{isl}, I)$ through a point contact when driven with a current I, for negligible phonon or photon heating. Blue indicates cooling of the island in Fig. 2(a), while red indicates heating. The solid curve is the steady state ($J = 0$), with the catastrophe at I_c. The straight line is the maximum current, I_{max}, corresponding to infinite bias.

Equation (2) is derived[1] by combining linear response terms (the Peltier effect due to the current I, and heat flow due to the temperature difference, $T_0 - T_{isl}$), with a nonlinear I^2 term corresponding to joule heating. Heat flow out of the island in Fig. 2(a), due to a current I passing though element 1, then the island, and then element 2, is

$$J(T_{isl}, I) \simeq \Pi_- I - \Theta_+(T_0 - T_{isl}) - \tfrac{1}{2}G_+^{-1} I^2, \qquad (3)$$

for $\Pi_- = \Pi_2 - \Pi_1$, $\Theta_+ = \Theta_1 + \Theta_2 + \Theta_{ph}$, and $G_+^{-1} = G_1^{-1} + G_2^{-1}$. Here Π_i, G_i, and Θ_i are the Peltier coefficient, and the electrical and thermal conductances of element i. The steady-state curve, $J = 0$, gives T_{isl} as a quadratic function of I. The parabola's minimum is T_{min} in Eq. (2) with $ZT = G_+\Pi_-^2/(\Theta_+ T)$.

For a point contact at pinch off (see Sec. III), linear response[28,32–37] gives $G_1 = (e^2/h)(1/2)$, $\Pi_1 = -(k_B T_0/e) 2\ln(2)$, and[38] $\Theta_1 = (k_B^2 T_0/h)(\pi^2/6 - 2[\ln(2)]^2)$. Thus $ZT \simeq 1.4$ so

$$\eta = 0.22(1 - T_0/T_{isl}), \quad T_{min} = 0.3T_0. \qquad (4)$$

However, Eq. (3) ceases to apply whenever thermoelectric effects are strong enough that the nonlinear terms that were not included in Eq. (3) become relevant. Heuristically, Eq. (3) fails to get the physics *qualitively* correct for any system where the nonlinear Peltier term,[21] $\widetilde{\Pi}\, I^2$, is larger than the joule heating term $\frac{1}{2}G_+^{-1} I^2$. The reason being that including $\widetilde{\Pi}\, I^2$ in Eq. (3) then changes the sign of the prefactor on I^2; so one must go beyond I^2 to find the steady-state curve's minimum. Including this $\widetilde{\Pi}$ term will then make the refrigerator *better* than if it were neglected. However, higher order terms (I^3 or higher) will then also be crucial in determining the lowest temperature the refrigerator can achieve.

This breakdown of the nearly linear theory as $\widetilde{\Pi}$ increases is discussed in Sec. VI for a particular system (a point contact in parallel with a tunnel barrier) in which $\widetilde{\Pi}$ can be varied; it indeed occurs when $\widetilde{\Pi}$ is of order $\frac{1}{2}G_+^{-1}$. Readers familiar with simple ϕ^4 theory will see this is similar to the para- to

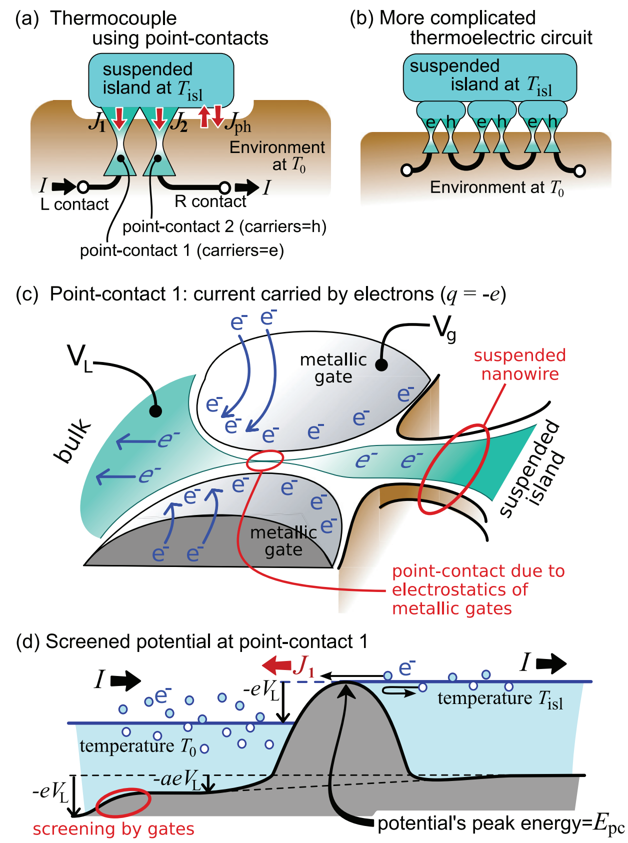

FIG. 2. (Color online) Thermoelectric circuits made with point contacts shown in (a), (b); "e" ("h") means the point contact is in a material whose charge carriers are electrons (holes). One should minimize the heat current carried by phonons and photons, J_{ph}, by suspending the island.[29,30] The temperature of the island in similar setups (albeit not suspended) has been probed experimentally using a quantum dot as thermometer,[31] although not yet in the regime of refrigeration. (c) Motion of charges in the gates (arrows) caused by making V_L positive, which partially screens V_L at some distance from the point contact. (d) Point contact 1 tuned to pinch off (E_{pc} equals the island's chemical potential) by adjusting V_g.

ferromagnetic crossover of a magnet in a B field upon reducing the temperature.[39] However, unlike in ϕ^4 theory, for small barrier transmission (or a point contact alone), the analysis shows such strong nonlinearities (catastrophe, etc.) that the minimum is not captured by a perturbative expansion in I up to any order.

Of course, the above heuristic argument assumes that nonlinearities in the thermoelectric response are significant when the linear thermoelectric response is significant. While in most cases this is true, the S-N tunnel junction is a counterexample; it has no thermoelectric response in the linear regime (so $ZT = 0$), but does have a large nonlinear response which has been used for refrigeration.[13–15] This is because the electron and holes have the same transmission at zero bias (so there is no thermoelectric response), but nonlinear charging effects enhance the transmission of electrons over those of holes creating an entirely nonlinear thermoelectric effect. It would be interesting to see if such S-N junctions exhibit the type of catastrophes found in this work for point contacts.

III. FULLY NONLINEAR ANALYSIS FOR POINT CONTACTS

The thermoelectricity literature[1–3] discusses $J(T_{\rm isl},I)$—as in Eq. (3) above—rather than $J(T_{\rm isl},V)$ for voltage drop, V. This is because different thermoelectric devices are arranged in series electrically [see Figs. 2(a) and 2(b)], so I is the same in all of them (unlike voltage drops). Thus it is easier to get response of a series of elements from each element's $J(T_{\rm isl},I)$ than from each element's $J(T_{\rm isl},V)$. For complicated nonlinear responses, the former is straightforward, while the latter is extremely difficult; thus I consider $J(T_{\rm isl},I)$.

I take the island to be classical; i.e., big enough for particles entering it to thermalize to a Fermi distribution at temperature $T_{\rm isl}$ before escaping. I also assume quantum charging effects (Coulomb blockade, etc.) within the island are negligible, while classical charging effects ensure electroneutrality (i.e., that the sum of electrical currents into the island is zero). References 40–45 consider cases where there is a quantum dot in place of the classical island. In our case, each point contact can be treated by a separate Landauer-Büttiker scattering matrix analysis;[28] see also Refs. 32–37. I generalize these heat currents to the nonlinear regime,[46] including electrostatic (Hartree-like) interaction effects in a self-consistent and gauge-invariant manner, as Refs. 24–27 did for charge current; see also Refs. 47 and 48. To go beyond the voltage-squared contributions to transport (which Ref. 24 treated in detail), I use a simple model of interactions, which is none the less gauge invariant and self consistent. The charge current, I_i, and heat current, J_i, into lead i of a given nanostructure are

$$I_i = -\int_{-\infty}^{\infty} \frac{d\epsilon}{h} \sum_j q\mathcal{A}_{ij}(\{\epsilon - qV_k\})\, f_j(\epsilon), \tag{5}$$

$$J_i = -\int_{-\infty}^{\infty} \frac{d\epsilon}{h} \sum_j (\epsilon - qV_i)\, \mathcal{A}_{ij}(\{\epsilon - qV_k\})\, f_j(\epsilon), \tag{6}$$

where $f_j(\epsilon) = (1+\exp[(\epsilon - qV_j)/(k_{\rm B}T_j)])^{-1}$ is the Fermi function, and q is the charge of the carriers; electrons with $q=-e$ in point-contact 1 and holes with $q=e$ in point-contact 2. The energy ϵ and all voltages V_k are measured from the same external reference. The transmission function of a particle through the nanostructure from lead j to lead i is $\mathcal{A}_{ij}(\{\epsilon - qV_k\}) = {\rm Tr}[\mathbf{1}_i\delta_{ij} - \mathcal{S}_{ij}^\dagger(\{\epsilon - qV_k\})\mathcal{S}_{ij}(\{\epsilon - qV_k\})]$, where $\mathcal{S}_{ij}$ is the scattering matrix from lead j to lead i, and the trace is over all modes of those leads. Here $\mathcal{S}_{ij}$ must be found *self-consistently*; it depends on the charge distribution in the nanostructure, which in turn depends on $\mathcal{S}_{ij}$. Writing $\mathcal{S}_{ij}$ as a function only of energy differences, $\{\epsilon - qV_k\}$, makes the gauge-invariance explicit; it satisfies[24] $[(d/d\epsilon)+\sum_k(d/d(qV_k))]\mathcal{A}_{ij}=0$.

Point-contact 1 is a two-lead nanostructure with electron charge carriers ($q=-e$). The gauge invariance means one is free to measure all energies ϵ and voltages V_k (including $V_{\rm g}$) from the island's chemical potential (the point-contact's M lead). I assume that a proportion $(1-a)$ of $V_{\rm L}$ is screened by the electrostatic gates a long way from the narrowest point of the point contact, while the rest is screened self-consistently by the electron gas [Fig. 2(d)] close to the point contact. Then the *screened* point contact induces a potential barrier of height, $E_{\rm pc}$ (measured from the island's chemical potential), typically obeying $E_{\rm pc} - E_{\rm g} = E_{\rm scr}(aqV_{\rm L})$, where $E_{\rm scr}$ is due to screening. Here $E_{\rm g}$ can be tuned at will, since it is $eV_{\rm g}$ minus a geometry-dependent constant. Assuming a long enough point contact that there is negligible tunneling, one has $\mathcal{A}_{\rm LM}(\epsilon - E_{\rm pc} > 0) = -1$ (perfect transmission) and $\mathcal{A}_{\rm LM}(\epsilon - E_{\rm pc} < 0) = 0$ (no transmission).[49] As an example, the Appendix gives a simple screening model for which I derive $E_{\rm scr}(aqV_{\rm L})$ self-consistently. However, in what follows I allow the nature of screening [both a and the form of $E_{\rm scr}(aqV_{\rm L})$] to be completely arbitrary.

For any given $V_{\rm L}$, one can adjust $V_{\rm g}$ to tune to pinch off ($E_{\rm pc}=0$). If the gates dominate screening ($a\to 0$), then $E_{\rm pc}$ is $V_{\rm L}$ independent, making this straightforward. Otherwise, the point contact should be calibrated prior to use; finding the pinch-off point (the $V_{\rm g}$ at which current starts to flow), as a function of $V_{\rm L}$. At pinch off, the currents from point-contact 1 into the island are

$$I(T_{\rm isl},V_{\rm L}) = \frac{ek_{\rm B}}{h}[T_{\rm isl}\ln(2) - T_0\ln(1+e^{-eV_{\rm L}/k_{\rm B}T_0})], \tag{7}$$

$$J_1(T_{\rm isl},V_{\rm L}) = -\frac{k_{\rm B}^2}{h}\left[T_{\rm isl}^2\frac{\pi^2}{12} + T_0^2{\rm Li}_2(-e^{-eV_{\rm L}/k_{\rm B}T_0})\right], \tag{8}$$

where ${\rm Li}_2(z)$ is a dilogarithm function. Equations (7) and (8) give

$$J_1(T_{\rm isl},I) = -\frac{k_{\rm B}^2T_0^2}{h}\left[\frac{\pi^2T_{\rm isl}^2}{12T_0^2} + {\rm Li}_2(1-\exp[\mathcal{I}(T_{\rm isl},I)])\right], \tag{9}$$

where I define $\mathcal{I} = h[I_{\max}(T_{\rm isl}) - I]/(ek_{\rm B}T_0)$ and note that $I \leqslant I^{\max}(T_{\rm isl}) = ek_{\rm B}T_{\rm isl}\ln[2]/h$. This function is given by the color plot in Fig. 1. For point-contact 2 (where carriers are holes not electrons) one takes $-e \leftrightarrow e$, then $J_2(T_{\rm isl},I) = J_1(T_{\rm isl},I)$ since $I_2 = -I$.

For $\mathcal{I} \ll 1$, one can use ${\rm Li}_2(z) = z + \mathcal{O}[z^2]$ to write

$$J_1 = (k_{\rm B}^2T_0^2/h)[\mathcal{I} - (\pi^2/12)(T_{\rm isl}/T_0)^2 + \mathcal{O}[\mathcal{I}^2]], \tag{10}$$

so J_1 as a quadratic in temperature and linear in current, the reverse of the nearly linear theory in Eq. (3). This approximation captures the features of the exact result plotted in Fig. 1, except the top-left corner. This corner is the linear-response regime [small $(T_0 - T_{\rm isl})$ and I], where one has Eq. (9) with ${\rm Li}_2(-1+z) \simeq -\pi^2/12 + \ln[2]z$.

IV. REFRIGERATION WITHOUT PHONONS OR PHOTONS

Heat flow into the island is $J_{\rm total} \propto J_1$ for the devices in Figs. 2(a) and 2(b); $J_{\rm total} = 2J_1$ for the thermocouple. The black curves in Fig. 1 are $J_{\rm total}=0$, giving the steady-state temperature (solid for stable steady states and dashed for unstable ones). Solid curves give the temperature the island will be cooled to by a current I. Equation (10) tells us the steady state has I as a quadratic function of $T_{\rm isl}$; this approximation gives the catastrophe at $eI_c/(ek_{\rm B}T_0) = 3(\ln[2]/\pi)^2 \simeq 0.14$ with $T_{\rm isl}/T_0 = 6\ln[2]/\pi^2 \simeq 0.42$, which is very close to the exact solution in Fig. 1.

A. Voltage dependence of cooling

As mentioned above, one typically considers the response of thermoelectric devices as a function of current I, rather

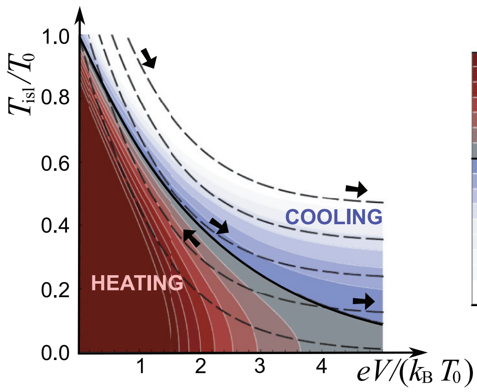

FIG. 3. (Color online) Heat current as a function of V and T_{isl}, with the steady state, $J = 0$, marked by the black curve [note the darkest color is used for all $hJ/(k_B T_0)^2 > 0.18$ and white for all $hJ/(k_B T_0)^2 < -0.25$]. Superimposed are lines of constant current (dashed); these are $hI/(ek_B T_0) = 0$, 0.08, 0.16, 0.24, 0.32 from bottom to top.

than voltage V, since I is conserved in series electrical circuits like those in Figs. 2(a) and 2(b). However, the origin of the catastrophe can be seen in Fig. 3, which compares the nonlinear response of the point contact as a function of voltage, Eq. (8), with curves of constant current given by Eq. (7). Curves with $I < I_c$, such as $hI/(ek_B T_0) = 0.08$, cross from cooling to heating and back again, while curves with $I > I_c$, such as $hI/(ek_B T_0) = 0.16, 0.24, 0.32$, never enter the heating regime. For larger I, the temperature saturates at a higher value. All of this fits with Fig. 1.

Note that if one wants the voltage response of the circuits in Figs. 2(a) and 2(b) one cannot easily get it from the voltage response of each element, as plotted in Fig. 3. This is because the voltage drop across each thermoelectric element depends nonlinearly on T_{isl}, even when the total voltage drop across all elements is fixed [unless all thermoelectric elements have the same $I(T_{isl}, V)$]. Indeed the simplest way to get the voltage response of the circuits in Figs. 2(a) and 2(b) is to take the heat flow in each circuit element *as a function of* I (as plotted in Fig. 1). Since I is the same in every element, one can get the heat flow as a function of the voltage drop across the circuit as the sum of voltage drops across each element (as a function of T_{isl} and I).

V. REFRIGERATION WITH PHONONS OR PHOTONS

I assume the metallic island is a suspended nanostructure[29] in a cryostat at 0.3 K, coupled to the substrate by suspended nanowires carrying the wires forming the cooling circuit. Wien's displacement law gives a photon wavelength of 10 mm at 0.3 K. For any island smaller than a few millimeters, bulk blackbody radiation is replaced by a single photon mode of noise flow between the island and its environment along the wires,[50,51] with $J_{ph} = r\alpha_0(T_0^2 - T_{isl}^2)$, where $\alpha_0 = \pi^2 k_B^2/(6h)$ is the "quantum" of heat flow and r is the mode's transmission. If the environment part of the circuit has an inductance[52] $> \mu$H (or its capacitor equivalent), then $r \ll 1$.

For a nanowire with N_{ph} phonon (vibrational) modes, $J_{ph} = \alpha(T_0^2 - T_{isl}^2)$, with $\alpha = N_{ph}\mathcal{T}\alpha_0$, with average transmission per mode of $\mathcal{T}$. Experimental nanowires[29] show $\alpha \sim 0.3\alpha_0$ at ambient temperature $T_0 = 0.3$ K. For this α, the steady-state curve has a pronounced cusp [uppermost v-shaped curve in Fig. 4(a)], very different from the parabola given by the standard nearly linear theory. Figure 4(b) shows that this cusp persists up to such larger α that there is very little refrigeration (note the vertical axis shows T_{isl}/T_0 is only slightly below one for any I). The α chosen for the plot in Fig. 4(b) corresponds to that observed in experiments in Refs. 29 and 30 at 3 K. This is about 30 times larger than the minimum phonon conductance observed in Ref. 29 at 0.3 K. It also shows that the nearly linear theory (dashed parabola) significantly underestimates the optimum refrigeration.

Taking the longitudinal phonon modes (velocity 9000 ms^{-1}) and three types of transverse modes (velocity 6000 ms^{-1}), the above experimental nanowires[30] (with cross section 200 nm$\times$100 nm) have $N_{ph} \sim 20$ at $T_0 \sim 0.3$ K. Evidently $\mathcal{T} \sim 1/60$; its smallness is probably due to the frequency mismatch between the phonons in the nanowire and the bulk. If wires with cross section 50 nm$\times$50 nm could be made, then $N_{ph} \sim 4$. Thus $\alpha \sim 0.06\alpha_0$ can be expected (i.e., five times smaller than the current nanowires); the dashed curve in Fig. 4(a) shows the catastrophe emerging at such α. To reduce α further, one can add surface roughness or serpentines.[30]

VI. CROSSOVER FOR POINT CONTACT IN PARALLEL WITH BARRIER

I ask how one can induce a transition to the parabolic behavior in Eq. (3), since phonons, etc., do not do so? I find

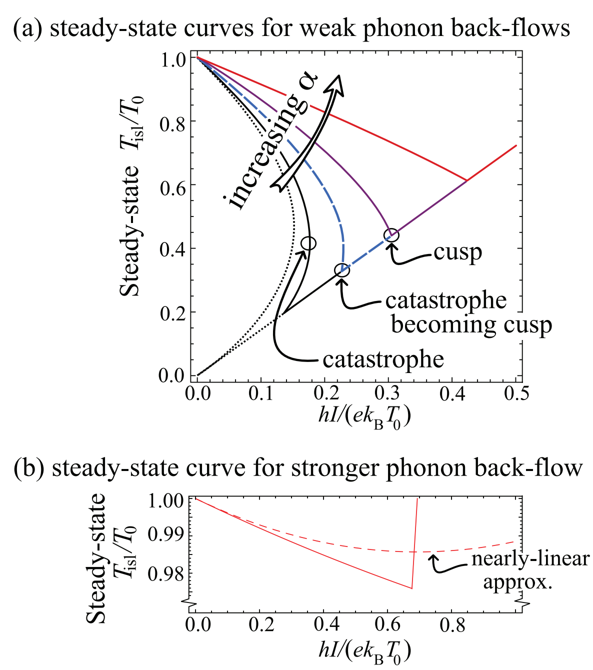

FIG. 4. (Color online) (a) Steady-state refrigeration curves, $J(T_{isl}, I) = 0$, for increasing heat flow due to phonons and photons; $\alpha/\alpha_0 = 0, 0.02, 0.06, 0.12, 0.3$. (b) The solid v-shaped curve is the steady-state refrigeration curve, $J(T_{isl}, I) = 0$ for $\alpha = 10\alpha_0$, i.e., about 30 times stronger phonon backflow than (a)'s uppermost curve. It is very different from the nearly linear theory for the same α (dashed parabola).

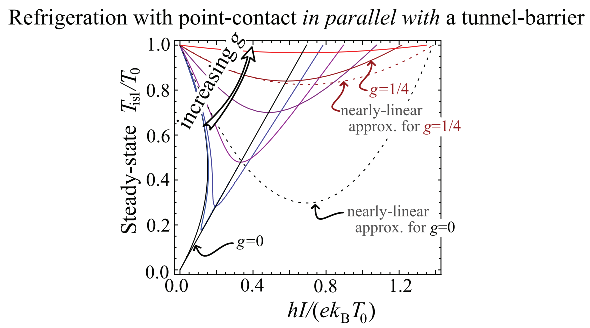

FIG. 5. (Color online) Steady-state refrigeration curves for a composite system; point contact in parallel with a barrier of conductance $G_{\mathrm{barrier}} = e^2 g/h$ for $g = 0,0.001,0.005,0.025,0.1,0.25,1$ (solid curves). The parabolas (dashed) are the nearly linear approximation for $g = 0,1/4$; for $g = 1$ the difference from the exact curve is not visible.

that a transition only occurs upon reducing the ϵ dependence of $\mathcal{A}_{ij}$, lowering the ratio of the thermoelectric response to the usual electric response—for example, replacing the point contact with a composite system consisting of a point contact in *parallel* with a tunnel barrier whose transmission is ϵ independent. Figure 5 shows the steady-state response of the composite system, for barrier conductance $G_{\mathrm{barrier}} = e^2 g/h$. Upon increasing g from zero, a transition occurs at $g = g_c \sim 1/200$; for $g > g_c$, the curve is single valued, so $T_{\min}$ becomes a continuous function of I. The nearly linear theory works for $g \gtrsim 1$; deviations are still visible for $g = 1/4$ (cf. solid and dashed curves). This fits the argument in Sec. II, since the composite system has $\widetilde{\Pi} < \frac{1}{2}G^{-1}$ for $g > (3\ln[2] - 1)/2 \simeq 0.53$.

VII. HEAT-ENGINE EFFICIENCY

Returning to the case of a point contact alone (without a tunnel barrier), I now consider its maximum efficiency as a heat engine. For $T_{isl} > T_0$, the circuit in Fig. 2(a) provides electrical power $P = IV$ to any load connected between L and R. To calculate the maximum electrical power $P(T_{isl}, I)$ that a heat engine can extract from a heat flow $J(T_{isl}, I)$, one assumes an ohmic load—so $V(T_{isl}, I) = I/G_{\mathrm{load}}$—is connected across its terminals, and adjust G_{load} to optimize the ratio of the power at the load $P(T_{isl}, I) = IV(T_{isl}, I)$ to the heat flow $J(T_{isl}, I)$. This corresponds to finding the $I = I_{\mathrm{opt}}$ which maximizes $P(T_{isl}, I)/J(T_{isl}, I)$. Maxima and minima are given by $P'J = PJ'$ where the primed is (d/dI). Given $V(T_{isl}, I)$ and $J(T_{isl}, I)$, one can solve this to find I_{opt}. Optimal efficiency is $\eta = P(I_{\mathrm{opt}})/J(I_{\mathrm{opt}})$.

As a warmup, I consider the usual linear problem, with $V(T_{isl}, I) = S(T_{isl} - T_0) - G^{-1}I$ and $J(T_{isl}, I) = \Theta(T_{isl} - T_0) + \Pi\, I$, with $\Pi = \Gamma/G$ and $S = B/G$. Calculating the optimal efficiency in the manner described above, I find $I_{\mathrm{opt}} = (\Theta/\Pi)[\sqrt{Z(T_{isl})T_{isl} + 1} - 1](T_{isl} - T_0)$. Dropping T dependences of ZT, one gets Eq. (1).

Now I use the same method to get the efficiency in the nonlinear regime. An analytic solution of $P'J = PJ'$ can be found for large (T_{isl}/T_0), using the fact (confirmed by the numerics) that in this limit $-eV_1(I_{\mathrm{opt}})/(k_BT_0) \gg 1$ with $V_1 < 0$. Otherwise the solution must be found numerically (see below).

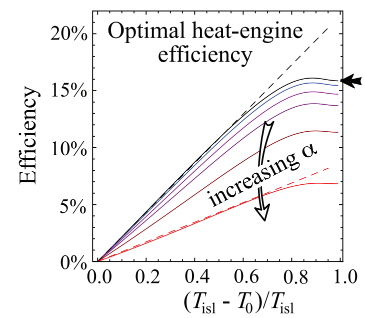

FIG. 6. (Color online) Heat-engine efficiency curves for $\alpha/\alpha_0 = 0,0.02,0.06,0.12,0.3,1$. Dashed lines are the linear-response predictions, Eq. (1), for $\alpha/\alpha_0 = 0,1$.

For large (T_{isl}/T_0), I take $\ln[1 + e^{\mu}] \to \mu$ and $\mathrm{Li}_2(-e^{\mu}) \to -\frac{1}{2}\mu^2$ for large μ, and find $eV_1(T_{isl}, I)/(k_BT_0) \simeq -t_{isl}\ln[2] + \tilde{I}$ with $hJ_1(T_{isl}, I)/(k_BT_0)^2 \simeq -\kappa t_{isl}^2/2 - \ln[2]\,\tilde{I}\,t_{isl} + \tilde{I}^2/2$, where I define $t_{isl} = T_{isl}/T_0$, $\tilde{I} = hI/(ek_BT_0)$, and $\kappa = \pi^2/6 - \ln^2[2] \simeq 1.16$. The heat current from the hot source into the device is $J(T_{isl}, I) = -J_1(T_{isl}, I)$, and $P = -V_1(T_{isl}, I)I$ (given that $V_1 < 0$). In this case $P'(T_{isl}, I_{\mathrm{opt}})J(T_{isl}, I_{\mathrm{opt}}) = P(T_{isl}, I_{\mathrm{opt}})J'(T_{isl}, I_{\mathrm{opt}})$ is a quadratic equation for I_{opt}; solving it gives

$$\frac{hI_{\mathrm{opt}}}{ek_BT_0} = \frac{\kappa}{\ln[2]}\left[\sqrt{1 + \ln^2[2]/\kappa} - 1\right]\frac{T_{isl}}{T_0}.$$

Without phonons or photons, the optimal efficiency tends to $1 - \sqrt{1 - 6(\ln[2]/\pi)^2} \simeq 15.9\%$ for $T_{isl} \to \infty$ (solid arrow in Fig. 6). Solving $P'J = PJ'$ with Eqs. (7) and (9) numerically, to find I_{opt} for different T_{isl}/T_0, I plot η against $(T_{isl} - T_0)/T_{isl}$ in Fig. 6. I have no simple argument why the curves are slightly peaked at $(T_{isl} - T_0)/T_{isl} \sim 0.85$.

VIII. CATASTROPHE AWAY FROM PINCH OFF

Figure 7 gives an example showing the catastrophe is still present when the point contact is away from pinch off; i.e., when the barrier's peak is above the chemical potential of the island in Fig. 1. The catastrophe is present when $0 < E_{pc} \lesssim k_BT_0$, which corresponds to the parameter regime where a significant thermoelectric response was found experimentally

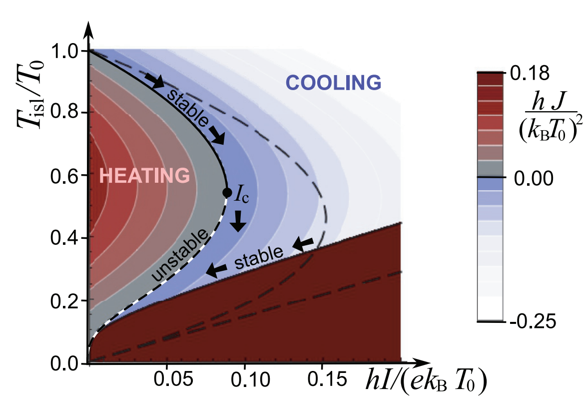

FIG. 7. (Color online) Heat current when the barrier peak is at $E_{pc} = k_BT_0/4$. The results which give this curve will be discussed elsewhere. For comparison, the thin-dashed curves are the steady state for $E_{pc} = 0$ in Fig. 1.

in Ref. 18. The formulas leading to Fig. 7 are similar to Eqs. (7)–(9) but rather longer, so I do not give them here.

IX. CONCLUDING REMARKS

I have shown that the point contact (arguably the simplest thermoelectric nanostructure) has a rich nonlinear behavior. In particular, when it is used as a refrigerator, it exhibits multiple steady states (stable and unstable) and a fold catastrophe, or a sharp cusp when there is significant phonon backflow. I see no reason to think that more complicated nanoscale thermoelectric systems[4–11] have less rich behaviors. Indeed a large ZT is a strong hint that its nonlinear Peltier term, $\widetilde{\Pi} I^2$ may dominate over its joule heating term $-\frac{1}{2}G^{-1}I^2$. Section II then gives a simple argument that ZT ceases to give even a qualitative indication of how good a refrigerator it is. Thus the fully nonlinear response of such systems requires detailed study, beyond the weak nonlinearities considered in Refs. 17, 21, 53, and 54.

Finally, I recall that this work considered the case where the charging effects of electrons at the point contact were well screened by the gates (or could be compensated for by the gates), meaning the $E_{\rm pc}$ in Fig. 2 does not significantly change with bias. Elsewhere, I will show that qualitatively similar effects can occur for point contacts (and other systems) without gates, for which $E_{\rm pc}$ depends on the bias.

Since the submission of this work a number of closely related works have appeared.[55–58]

ACKNOWLEDGMENTS

I thank F. Hekking, G. Rastelli, C. Stafford, and particularly Ph. Jacquod for insightful comments at various stages of this work.

APPENDIX: EXAMPLE OF A SELF-CONSISTENT SOLUTION

Here is a simple model of the point contact for which the self-consistent solution can be found easily. However, the results in the body of the manuscript apply for almost any self-consistent model. The point contact is treated as a one-dimensional scattering problem (along the x axis); see Fig. 2(d). Close to the point contact, this takes the form $qV(x) = E_{\rm g} - \kappa x^2 + qV_{\rm scr}(x - x_{\rm pc})$ with energy measured from the island's chemical potential. The transverse confinement induces the $(E_{\rm g} - \kappa x^2)$ term, where $E_{\rm g}$ can be tuned, since it equals $eV_{\rm g}$ minus a geometry-dependent constant. The $qV_{\rm scr}$ term is screening inside the electron gas, which I take as

$$V_{\rm scr}(x) = \begin{cases} aV_{\rm L} & \text{for} \quad x < -l_{\rm scr}, \\ aV_{\rm L}(l_{\rm scr} - x)/(2l_{\rm scr}) & \text{for} \quad |x| \leqslant l_{\rm scr}, \\ 0 & \text{for} \quad x > l_{\rm scr}, \end{cases}$$

with $x_{\rm pc}$ being the self-consistently determined peak of $qV(x)$. A little algebra gives $x_{\rm pc} = -aqV_{\rm L}/(4\kappa l_{\rm scr})$; thus the energy at the peak is $E_{\rm pc} = qV(x_{\rm pc}) = E_{\rm g} + \frac{1}{2}aqV_{\rm L}(1 - aqV_{\rm L}/(8\kappa l_{\rm scr}^2))$. Finally, I note that both a and $l_{\rm scr}$ depend on the scattering matrix of the junction, which in turn depends on $E_{\rm pc}$. To solve this problem self-consistently, I assume one is in the regime where $e_{\rm pc} = E_{\rm pc} - E_{\rm g}$ is small enough to approximate $a = a_0(1 + b_a e_{\rm pc})$ and $l_{\rm scr} = l_{\rm scr0}(1 + b_l e_{\rm pc})$. If necessary, $a_0, l_{\rm scr0}, b_a, b_l$ can be found by simulating Poisson's equation; typically $e_{\rm pc}$ is small for small a. Then $E_{\rm pc}$ is equal to a linear function of itself; rearranging this gives

$$E_{\rm pc} - E_{\rm g} = \frac{a_0 qV_{\rm L}/2 - C(qV_{\rm L})}{1 - a_0 b_a qV_{\rm L} + 2C(qV_{\rm L})[b_a - b_l]},$$

where I define $C(qV_{\rm L}) = (a_0 qV_{\rm L}/l_{\rm scr0})^2/(16\kappa)$. Thus the right-hand side of this equation is the $E_{\rm scr}(aqV_{\rm L})$ mentioned in the body of the text. As mentioned earlier, I assume that tunneling at energies $\epsilon < E_{\rm pc}$ is negligible, so $\mathcal{A}_{\rm LM}(\epsilon - E_{\rm pc} > 0) = -1$ and $\mathcal{A}_{\rm LM}(\epsilon - E_{\rm pc} < 0) = 0$. To see that this respects gauge invariance, I recall that $\epsilon, E_{\rm pc}, V_{\rm L,g}$ are all measured relative to the island's potential, and replace them by quantities measured from a fixed external reference, so the island is at $\widetilde{V}_{\rm M}$. For clarity, here [unlike in the paragraph containing Eq. (6)] it is necessary to use a tilde to explicitly indicate quantities measured from the external reference. I make the replacement $qV_{\rm L} = (\widetilde{\epsilon} - q\widetilde{V}_{\rm M}) - (\widetilde{\epsilon} - q\widetilde{V}_{\rm L})$. From this, one sees that $\mathcal{A}_{\rm LM}(\epsilon - E_{\rm pc})$ is only a function of the set of differences $\{\widetilde{\epsilon} - q\widetilde{V}_k\}$, and so respects gauge invariance.

[1] H. J. Goldsmid, *Thermoelectric Refrigeration* (Temple Press, London, 1964), Chap. 1; *Introduction to Thermoelectricity* (Springer, Heidelberg, 2009), Chap. 2.
[2] F. J. DiSalvo, Science **285**, 703 (1999).
[3] A. Shakouri and M. Zebarjadi, in *Thermal Nanosystems and Nanomaterials*, edited by S. Volz (Springer, Heidelberg, 2009), Chap. 9; A. Shakouri, Annu. Rev. Mater. Res. **41**, 399 (2011).
[4] G. Casati, C. Mejía-Monasterio, and T. Prosen, Phys. Rev. Lett. **101**, 016601 (2008).
[5] D. Nozaki, H. Sevinçli, W. Li, R. Gutiérrez, and G. Cuniberti, Phys. Rev. B **81**, 235406 (2010).
[6] K. K. Saha, T. Markussen, K. S. Thygesen, and B. K. Nikolić, Phys. Rev. B **84**, 041412(R) (2011).
[7] M. Wierzbicki and R. Swirkowicz, Phys. Rev. B **84**, 075410 (2011).
[8] O. Karlström, H. Linke, G. Karlström, and A. Wacker, Phys. Rev. B **84**, 113415 (2011).
[9] T. Gunst, T. Markussen, A.-P. Jauho, and M. Brandbyge, Phys. Rev. B **84**, 155449 (2011).
[10] G. Rajput and K. C. Sharma, J. Appl. Phys. **110**, 113723 (2011).
[11] P. Trocha and J. Barnaś, Phys. Rev. B **85**, 085408 (2012).
[12] F. Giazotto, T. T. Heikkila, A. Luukanen, A. M. Savin, and J. P. Pekola, Rev. Mod. Phys. **78**, 217 (2006); J. T. Muhonen, M. Meschke, and J. P. Pekola, Rep. Prog. Phys. **75**, 046501 (2012).
[13] M. M. Leivo, J. P. Pekola, and D. V. Averin, Appl. Phys. Lett. **68**, 1996 (1996).

[14]A. M. Clark, N. A. Miller, A. Williams, S. T. Ruggiero, G. C. Hilton, L. R. Vale, J. A. Beall, K. D. Irwin, and J. N. Ullom, Appl. Phys. Lett. **86**, 173508 (2005).

[15]S. Rajauria, P. S. Luo, T. Fournier, F. W. J. Hekking, H. Courtois, and B. Pannetier, Phys. Rev. Lett. **99**, 047004 (2007); S. Rajauria, P. Gandit, F. W. J. Hekking, B. Pannetier, and H. Courtois, J. Low Temp. Phys. **154**, 211 (2009).

[16]For bulk semiconductors, linear response (Boltzmann) transport theory works when the temperature difference on the scale of the inelastic scattering length is small compared with the average temperature at that point. The inelastic scattering length (typically 1–100 nm at 290 K) is very much less than the length over which temperature drops (length of semiconductor, e.g., millimeters). Then a linear theory works well even when the temperature drop is large, although nonlinear corrections have been studied.[17] In contrast, nanostructures are often smaller than the inelastic scattering length (often much more than 1 μm at 1 K). Ideally the whole temperature drop occurs across this nanostructure, making linear-response theory inappropriate when the temperature drop is significant.

[17]M. Zebarjadi, K. Esfarjani, and A. Shakouri, Appl. Phys. Lett. **91**, 122104 (2007); MRS Proc. **1044**, U10-04 (2007).

[18]L. W. Molenkamp, Th. Gravier, H. van Houten, O. J. A. Buijk, M. A. A. Mabesoone, and C. T. Foxon, Phys. Rev. Lett. **68**, 3765 (1992); H. van Houten, L. W. Molenkamp, C. W. J. Beenakker, and C. T. Foxon, Semicond. Sci. Technol. **7**, B215 (1992).

[19]U. Ghoshal, S. Ghoshal, C. McDowell, L. Shi, S. Cordes, and M. Farinelli, Appl. Phys. Lett. **80**, 3006 (2002).

[20]E. N. Bogachek, A. G. Scherbakov, and U. Landman, Solid State Commun. **108**, 851 (1998); this work calculates the nonlinear differential Peltier coefficient $\Pi_{\text{diff}} = dJ/dI$, even though some formulas assume linear behavior.

[21]M. A. Çipiloğlu, S. Turgut, and M. Tomak, Phys. Status Solidi B **241**, 2575 (2004).

[22]N. Nakpathomkun, H. Q. Xu, and H. Linke, Phys. Rev. B **82**, 235428 (2010).

[23]M. V. Moskalets, JETP Lett. **62**, 719 (1995).

[24]T. Christen and M. Büttiker, Europhys. Lett. **35**, 523 (1996).

[25]D. Sanchez and M. Buttiker, Phys. Rev. Lett. **93**, 106802 (2004).

[26]J. Meair and Ph. Jacquod, J. Phys.: Condens. Matter **24**, 272201 (2012).

[27]D. Sánchez and R. López, Phys. Rev. Lett. **110**, 026804 (2013).

[28]P. N. Butcher, J. Phys.: Condens. Matter **2**, 4869 (1990).

[29]J. S. Heron, T. Fournier, N. Mingo, and O. Bourgeois, Nano Lett. **9**, 1861 (2009).

[30]J.-S. Heron, C. Bera, T. Fournier, N. Mingo, and O. Bourgeois, Phys. Rev. B **82**, 155458 (2010).

[31]S. Gasparinetti, F. Deon, G. Biasiol, L. Sorba, F. Beltram, and F. Giazotto, Phys. Rev. B **83**, 201306(R) (2011).

[32]H.-L. Engquist and P. W. Anderson, Phys. Rev. B **24**, 1151 (1981).

[33]U. Sivan and Y. Imry, Phys. Rev. B **33**, 551 (1986).

[34]N. R. Claughton and C. J. Lambert, Phys. Rev. B **53**, 6605 (1996).

[35]Ph. Jacquod and R. S. Whitney, Europhys. Lett. **91**, 67009 (2010).

[36]Y.-S. Liu, B. C. Hsu, and Y. C. Chen, J. Phys. Chem. C **115**, 6111 (2011).

[37]Ph. Jacquod, R. S. Whitney, J. Meair, and M. Büttiker, Phys. Rev. B **86**, 155118 (2012).

[38]Θ_i is the thermal conductance for $I = 0$ (not $V = 0$).

[39]In ϕ^4 theory at high temperatures, the ϕ^2 term is positive and determines the minimum of free energy. At low temperatures, the ϕ^2 term is negative and the minimum is determined by a higher order term (the ϕ^4 term). In the refrigerator, the role of ϕ is played by I, while the parameter which controls the sign of the quadratic term is $\widetilde{\Pi}$ rather than temperature.

[40]O. Entin-Wohlman, Y. Imry, and A. Aharony, Phys. Rev. B **82**, 115314 (2010).

[41]B. Sothmann, R. Sánchez, A. N. Jordan, and M. Büttiker, Phys. Rev. B **85**, 205301 (2012).

[42]B. Sothmann and M. Büttiker, Europhys. Lett. **99**, 27001 (2012).

[43]R. Sánchez and M. Büttiker, Europhys. Lett. **100**, 47008 (2012).

[44]J.-H. Jiang, O. Entin-Wohlman, and Y. Imry, New J. Phys. **15**, 075021 (2013).

[45]O. Entin-Wohlman, A. Aharony, and Y. Imry, arXiv:1306.1813.

[46]R. S. Whitney, Phys. Rev. B **87**, 115404 (2013).

[47]M. Büttiker, J. Phys.: Condens. Matter **5**, 9361 (1993); M. Büttiker, A. Prêtre, and H. Thomas, Phys. Rev. Lett. **70**, 4114 (1993); Z. Phys. B **94**, 133 (1994); T. Christen and M. Büttiker, *ibid.* **77**, 143 (1996).

[48]C. Petitjean, D. Waltner, J. Kuipers, I. Adagideli, and K. Richter, Phys. Rev. B **80**, 115310 (2009).

[49]The results are unchanged if a little bit of tunneling means that $\mathcal{A}_{\text{LM}}(\epsilon)$ goes smoothly from 0 to -1 on a scale less than $k_{\text{B}}T_{\text{isl}}$.

[50]J. B. Pendry, J. Phys. A: Math. Gen. **16**, 2161 (1983).

[51]D. R. Schmidt, R. J. Schoelkopf, and A. N. Cleland, Phys. Rev. Lett. **93**, 045901 (2004).

[52]L. M. A. Pascal, H. Courtois, and F. W. J. Hekking, Phys. Rev. B **83**, 125113 (2011).

[53]J. Meair and Ph. Jacquod, J. Phys.: Condens. Matter **25**, 082201 (2013).

[54]R. Lopez and D. Sanchez, Phys. Rev. B **88**, 045129 (2013).

[55]A. N. Jordan, B. Sothmann, R. Sánchez, and M. Büttiker, Phys. Rev. B **87**, 075312 (2013).

[56]S.-Y. Hwang, D. Sánchez, M. Lee, and R. López, arXiv:1306.6558.

[57]R. S. Whitney, arXiv:1306.0826.

[58]S. Hershfield, K. A. Muttalib, and B. J. Nartowt, arXiv:1307.5670.